Predicting the speech reception threshold of cochlear implant listeners using an envelope-correlation based measure

Nima Yousefian^{a)} and Philipos C. Loizou^{b)}

Department of Electrical Engineering, University of Texas at Dallas, Richardson, Texas 75080

(Received 19 January 2012; revised 21 August 2012; accepted 30 August 2012)

A modulation-based index is proposed for predicting speech intelligibility by cochlear implant (CI) listeners. The input to the proposed index are speech envelopes extracted using the individual CI user's daily strategy, and as such, this approach incorporates information about the number of active electrodes, shape of the compression function and electrical dynamic range. High correlation ($r = 0.96$) was achieved with the proposed index when evaluated with speech-reception thresholds (SRTs) obtained by CI users in steady and speech-masker conditions. This outcome suggests that the information contained in electrograms seems to be sufficient for reliably predicting CI user's performance in noise. The proposed index can be used by clinicians to optimize the selection of fitting parameters of individual CI users for better performance in noise.

© 2012 Acoustical Society of America. [http://dx.doi.org/10.1121/1.4754539]

PACS number(s): 43.66.Ts, 43.71.Ky [EB]

Pages: 3399–3405

I. INTRODUCTION

A number of indices, such as the speech-transmission index (STI) (Houtgast and Steeneken, 1985) and speech intelligibility index (SII) (ANSI, 1997), have been proposed for predicting the intelligibility of speech in noise. These indices have been evaluated and validated for the most part using data collected with normal-hearing (NH) and hearing-impaired listeners, rather than with CI users. Traditional indices (e.g., SII) assume critical-band spectral representations which are intended for modeling normal auditory frequency selectivity. Furthermore, these indices assume that the SNR computed in each band and limited within the range of $[-15 \text{ dB}, +15 \text{ dB}]$, are linearly mapped to the range of $[0, 1]$, with 1 indicating the highest index value (hence, the highest intelligibility score) and 0 the lowest index value. That is, it is assumed that intelligibility varies linearly, within a 30-dB SNR dynamic range, as the band SNR increases. This assumption has been found to work well with NH listeners, but has never been validated with CI users. For one, the linear segment of the CI users' psychometric function is much narrower and spans the range of 0 to 12 dB (Friesen *et al.*, 2001), with performance reaching floor for negative SNR values. In brief, some of the assumptions made when computing the SII and STI indices do not hold when speech is processed via cochlear implants.

The present study relaxes the above assumptions and evaluates the predictive power of a modified speech-based STI measure. An STI-like measure is chosen as it relies on information present in envelope modulations, similar to that presented to CI users with existing coding strategies. This measure has been found previously to correlate highly with vocoded speech perception (Chen and Loizou, 2011), and is adapted in the present study for cochlear implants to include

information such as limited dynamic range and number of active electrodes (acoustic frequency resolution), among others. More precisely, the input to this measure are the speech envelopes extracted using the individual CI user's MAP after implant processing. This input reflects the shape of loudness growth function (compression function), thresholds and most comfortable levels (i.e., electrical dynamic range) and number of active electrodes (frequency resolution). The output of the proposed measure is correlated with intelligibility scores (measured in terms of SRT) obtained by CI users in a number of conditions involving steady and competing-talker backgrounds.

The aim of the present study is to develop an intelligibility index that can be used to predict the SRT of individual CI users using information readily available to the audiologist, i.e., the user's MAP. This index will understandably have its limitations as it will not include neuropathologically related information (e.g., duration of deafness) that might also affect the performance of CI users. Such an index could benefit the audiologists in different ways. It could, for instance, help the audiologist optimize the choice of specific fitting parameters for better performance in noisy conditions. Also, if given the choice of selecting and programming the user with one or two "noise" programs, the index could be used to select the right "noise" program for a particular CI user. Finally, the intelligibility index could be used by engineers who design noise reduction algorithms to optimize and improve their algorithms.

II. METHODS

A. Subjects and conditions

A total of ten postlingually deafened CI patients (mean age = 63 yr) participated in the experiment. The participants were all fitted with the Nucleus multichannel implant device manufactured by Cochlear, Ltd. Biographical data for all subjects are presented in Table I. The speech stimuli used for testing were sentences taken from the IEEE corpus

^{a)} Author to whom correspondence should be addressed. Electronic mail: nimayou@utdallas.edu

^{b)} Deceased.

TABLE I. Biographical data and fitting parameters of the 10 subjects tested.

Subject	Age	Gender	Years implanted (L/R)	Number of active electrodes	Stimulation rate (Hz)	Mean dynamic range (L/R) (dB)	Q Value	CI processor	Etiology of hearing loss
S1	48	F	8/8	22	800	8.71/8.34	20	ESPril 3G	Unknown
S2	72	F	9/9	22	889	7.07/6.81	20	Freedom	Unknown
S3	77	M	6/6	22	889	2.57/2.47	20	Freedom	Hereditary
S4	63	F	8/8	20	889	8.15/7.25	20	ESPril 3G	Rubella
S5	56	F	4/7	22	889	1.66/6.81	20	Freedom	Unknown
S6	70	F	6/-	22	1231	5.01/7.25	20	Freedom	Unknown
S7	52	F	2/1	22	889	3.80/6.45	20	Nucleus 5	Unknown
S8	65	F	15/8	22	1143	8.76/7.59	20	Freedom	Unknown
S9	60	F	6/9	22	1455	10.45/8.48	20	Freedom	Hereditary
S10	70	F	3/1	20	889	8.71/5.91	20	Nucleus 5	Unknown

(IEEE, 1969). Each sentence was composed of 7–12 words and was produced by a single male talker. Two types of noise (speech-shaped and competing talker) were used as maskers. The spectrum of the speech-shaped noise (SSN) matched the average long-term spectrum of the speech materials. The competing talker (CT) sentences were taken from the AzBio corpus (Spahr *et al.*, 2012). This corpus contains sentences uttered by two female and two male speakers.

Since the prevalent noise-reduction algorithm available commercially today in cochlear implants (and hearing aids) makes use of two closely spaced microphones, a comparison is made in this study of three different dual-mic algorithms. Although the algorithms take input from two microphones, they produce a single (enhanced) output that is fed into the CI user's processor. The algorithms are tested in standard spatial configurations where the simulated target location is always assumed to be at the front (0° azimuth). Listeners were tested in conditions with either one or two noise sources were present at various simulated spatial positions, and in some cases originating from both hemifields. Table II summarizes the experimental conditions. The noisy stimuli were processed using the following dual-microphone noise reduction methods: (1) a software directional microphone with hypercardioid polar diagram, (2) an adaptive beam-former (Berge and Wouters, 1998), and (3) an extension of the coherence-based approach proposed by Yousefian and Loizou (2012). In the rest of the paper, these processing methods are referred to as method A, B, and C respectively. It should be pointed out that Methods A and B are commercially available in Nucleus 5 (Cochlear, Ltd.). In total, there were 12 conditions (3 methods $\times$ 2 spatial configurations $\times$ 2 noise types) for the single interferer scenario, and six different conditions (3 methods $\times$ 2 spatial configurations) for multiple-interferer scenarios. The stimuli were streamed directly to the CI users at a comfortable level using the Cochlear Ltd.'s Nucleus Implant CommunicatorTM (NIC). All stimuli were presented to the subjects unilaterally (for bilateral users, the ear with the highest score was used).

B. SRT measurements

SRT measurements were made using a method similar to that by Hawley *et al.* (2004). At the start of each SRT

measurement, the subjects were presented with the stimuli at low SNR levels. After their response, the stimulus was replayed with +4 dB increase in SNR, until the subjects reproduced more than half of the sentence correctly. From that point forward, the SRT measurement began using a one-down/one-up adaptive SRT technique targeting 50% correct speech reception (Levitt, 1971). Each IEEE sentence had five designated keywords. The SNR of the next trial was raised by 2 dB if two or fewer keywords were correct and decreased by 2 dB if three or more keywords were correct. In this work, SRT was determined by averaging the SNR level presented in the last eight trials (trials 4 through, and including, 11). Figure 1 shows the mean SRT values obtained in various conditions by the ten CI users (Yousefian and Loizou, 2011). Speech perception in quiet of all ten subjects were above 50%. In addition, in our experiments the SRT converged for all subjects and noise conditions, where the highest SNR corresponds to SRT level was +12 dB.

III. MODEL DESCRIPTION

The proposed envelope-correlation based measure (ECM) is similar to the normalized covariance measure (NCM) (Holube and Kollmeier, 1996) evaluated by Goldsworthy and Greenberg (2004). The NCM is a speech-based STI measure and has been shown previously to yield high correlation with vocoded speech perception (Chen and Loizou, 2011) and speech subjected to non-linear distortions, such as those introduced by spectral-subtractive algorithms (Ma *et al.*, 2009). This measure computes the STI as a weighted sum of transmission index (TI) values determined from the envelopes of the probe (clean speech) and response (processed or corrupted) signals in each frequency band. The difference between the traditional STI and NCM is that the STI quantifies the change in modulation depth between the probe and response¹ envelopes using the modulation transfer function, while the NCM measure is based on the covariance (Pearson correlation) between these two envelopes.

A. Envelope-correlation based measure (ECM)

The input to the ECM index are the speech envelopes obtained using the advanced combination encoder (ACE) strategy (Vandali *et al.*, 2000) programmed with the individual

TABLE II. List of spatial configurations tested.

No. of interferers	Interferer type	Angle of incidence
One interferer	speech-shaped noise	45°
One interferer	female talker	45°
One interferer	speech-shaped noise	90°
One interferer	female talker	90°
Two interferers	female and male talkers	90°, 180°
Two interferers	female and male talkers	90°, -90°

CI user's settings (MAP). That is, the input to the ECM index are the electrodiagrams of the noise-free (clean) speech (used as a reference) and the noisy stimuli (i.e., same speech, plus noise) processed by one of the noise reduction methods. As such, the ECM index incorporates information about the shape of the compression function (Q value), number of active electrodes, and thresholds and comfortable levels (electrical dynamic range). The ACE implementation was identical to that used in the NIC (Cochlear, Ltd.) system. The ECM model does not explicitly use the actual values of stimulation rate (e.g., 900 pps), dynamic range (e.g., 10 dB), etc. Nearly all information contained in the CI user's MAP is implicitly contained in the envelopes (with the exception of pulse width, electrode configuration, etc). The differences in stimulation rate, for instance, are reflected by the degree with which temporal details or modulations are represented in the envelopes, with high stimulation rates preserving fine temporal details (high modulation frequencies) and low rates capturing only gross temporal details (low modulation frequencies). For intelligibility prediction, however, only low-frequency envelope modulations are required (e.g., Houtgast and Steeneken, 1985).

Let $x_k(t)$ be the envelope (after compression) in band k of the noise-free (clean) speech obtained after implant processing (The ACE strategy), and $y_k(t)$ that of the speech stimuli in noise processed by noise-reduction algorithms,

obtained after the same implant processing. These envelopes are subsequently downsampled from stimulation rate to 50 Hz limiting the envelope modulations to 0–25 Hz. This range was chosen as it encompasses the range (0–16 Hz) that several modulation filtering studies have shown to contribute the most to intelligibility (e.g., Drullman *et al.*, 1994; Elliott and Theunissen, 2009; Wójcicki and Loizou, 2012). The covariance between the compressed envelopes [denoted as $\hat{x}_k(t)$ and $\hat{y}_k(t)$] is computed as

$$r_k = \frac{\sum_t [\hat{x}_k(t) - \mu_x][\hat{y}_k(t) - \mu_y]}{\sqrt{\sum_t [\hat{x}_k(t) - \mu_x]^2} \sqrt{\sum_t [\hat{y}_k(t) - \mu_y]^2}}, \quad (1)$$

where μ_x and μ_y denote the mean of the clean and processed envelopes respectively in band k . The values of r_k in the above equation are bounded within $|r_k| \leq 1$, and indicate the degree that the clean speech (noise free) [i.e., $x_k(t)$] and processed (by a noise reduction processing method) [i.e., $y_k(t)$] signals are linearly related. The ECM index is finally computed as

$$\text{ECM} = \frac{1}{|\Delta|} \sum_{k \in \Delta} r_k^2, \quad (2)$$

where Δ indicates the set of active electrodes and $|\Delta|$ denotes the number of active electrodes.

Note that unlike the NCM index, the ECM index does not convert the squared correlations into an apparent SNR nor does it map the SNR to the range of 0 and 1. In the SII and STI implementations, for instance, when the computed SNR in a given band is 0 dB, it is mapped to SII = 0.5 or STI = 0.5 (prior to multiplication by the band-importance function) (ANSI, 1997). A value of SII = 0.5, however, is mapped to a 50%–100% intelligibility score depending on the speech material (sentences vs words) used (Fletcher and

FIG. 1. Mean SRT values obtained by ten CI users in the various conditions involving competing-talkers (CT) and speech-shaped noise (SSN). Error bars indicate standard deviations.

Galt, 1950). Clearly, such a mapping would not be suitable for cochlear implant patients who generally perform poorly, near 0% correct (Friesen *et al.*, 2001), at an input SNR of 0 dB. Also, as shown in Eq. (2), the ECM does not use a band-importance function, as done in SII (ANSI, 1997), since the band-importance functions were derived and validated with NH listeners.

B. Determining the ECM value corresponding to SRT

Determination of the SRT based on ECM values involved two stages: a training stage and a testing stage. During the training stage, the ECM value corresponding to the SRT values, i.e., 50% correct score, of a subset of CI users was determined. More precisely, the ECM values for three subjects (S3, S6, and S7, selected randomly) at the SNR levels corresponding to their observed SRTs were computed. For example, let us assume that the observed SRT value for a given subject is 6.5 dB. Sentence stimuli were mixed with the masker at SNR = 6.5 dB, processed by the corresponding suppression algorithm and the ECM values were computed and averaged across all sentences used in that particular condition. The corresponding average ECM value was the ECM value corresponding to 50% correct and used to predict the SRT of the subject. This value was denoted as $ECM_{50\%}$. Including six noise conditions and three processing methods, a total of 18 $ECM_{50\%}$ values were obtained for each of the three subjects. Table III lists the $ECM_{50\%}$ values computed at the SNR levels corresponding to the observed SRTs. One IEEE sentence list (ten sentences) was used for the computation of the ECM values for each condition. By taking the average across the three subjects, $\overline{ECM}_{50\%} = 0.65$ is obtained. This $\overline{ECM}_{50\%}$ number is the ECM value corresponding to the average (across the three subjects) SRT level. In the present study, the $\overline{ECM}_{50\%}$ values (obtained by averaging $ECM_{50\%}$ values across

TABLE III. Individual $ECM_{50\%}$ values at the SRT in various conditions by three subjects (S3, S6, S7).

Method and Noise Condition	S3	S6	S7
Method A, SSN 45°	0.69	0.67	0.67
Method A, SSN 90°	0.74	0.71	0.71
Method A, CT 45°	0.55	0.61	0.66
Method A, CT 90°	0.71	0.71	0.74
Method A, CT (90°, 180°)	0.66	0.61	0.69
Method A, CT (90°, -90°)	0.70	0.67	0.72
Method B, SSN 45°	0.64	0.63	0.56
Method B, SSN 90°	0.59	0.63	0.55
Method B, CT 45°	0.64	0.63	0.59
Method B, CT 90°	0.66	0.56	0.59
Method B, CT (90°, 180°)	0.70	0.58	0.56
Method B, CT (90°, -90°)	0.70	0.66	0.66
Method C, SSN 45°	0.56	0.59	0.57
Method C, SSN 90°	0.78	0.77	0.68
Method C, CT 45°	0.56	0.55	0.65
Method C, CT 90°	0.77	0.77	0.68
Method C, CT (90°, 180°)	0.62	0.63	0.66
Method C, CT (90°, -90°)	0.54	0.65	0.60
Mean (SD)	0.66 (0.08)	0.65 (0.06)	0.64 (0.06)

subjects) were chosen because the intent was to predict group performance. Alternatively, one could compute the $\overline{ECM}_{50\%}$ value for each subject, and use that to predict individual CI user's performance.

IV. MODEL PREDICTION

In the testing stage, the SRT value for a given subject and condition can be predicted as follows. For a given condition, the input SNR is varied from -12 dB to 12 dB, in steps of 2 dB, and the ECM values of the processed stimuli are computed for each SNR level. The predicted SRT value is the SNR value corresponding to $\overline{ECM}_{50\%} = 0.65$. Linear interpolation is used in case the $\overline{ECM}_{50\%}$ value falls between two SNR levels. Figure 2 shows an example prediction of SRT for subject S2 for one condition. The measured SRT for this condition was 1.5 dB and the predicted SRT (corresponding to $\overline{ECM}_{50\%}$) was 1.04 dB. The results of the predictions with the proposed algorithm for each of the ten subjects are shown in Table IV. In the Table, the correlation coefficients and root-mean-square errors (RMSEs) are calculated based on the observed and predicted SRT values for all 18 noise conditions listed in Table III.

Figure 3 shows the scatter plot of observed SRT and predicted SRT values for all conditions tested. Mean SRT values obtained from all ten subjects were used in the analysis. The diagonal line indicates the points where the predicted and observed SRTs are equal to each other. The resulting correlation coefficient was $r = 0.96$ and the RMSE was 1.94 dB. Correlation analysis was also run using only the seven subjects that were not included in the training stage. The resulting correlation coefficient remained high at $r = 0.95$, while the RMSE increased to 2.25 dB. As shown in Fig. 3, and indicated by the small RMSE, the predicted SRT values are in good agreement with the observed SRT values. The data points in the figure are labeled according to both noise type (Fig. 3, top) and the processing method (Fig. 3, bottom).

FIG. 2. Predicted psychometric function for subject S2 based on data collected in one condition (method A, SSN 90°). The predicted SRT (corresponding to $ECM = 0.65$) was 1.04 dB (measured SRT was 1.5 dB). The horizontal line indicates $\overline{ECM}_{50\%} = 0.65$.

TABLE IV. Individual correlation and RMSE values for the ten subjects, considering all processing methods and noise conditions.

Subject	Correlation Coefficient	RMSE
S1	0.94	1.79
S2	0.97	1.53
S3	0.93	2.2
S4	0.70	5.24
S5	0.87	4.22
S6	0.96	1.55
S7	0.93	2.4
S8	0.92	3.59
S9	0.91	2.19
S10	0.95	2.03
Mean (SD)	0.91 (0.08)	2.67 (1.25)

The input to the proposed model was subject specific and accounted for information such as electrical dynamic range, shape of loudness growth function, and number of active electrodes. This allowed the model to capture to some degree the inherent variability in performance among CI

FIG. 3. Scatter plot of predicted SRT (via ECM) and observed SRT values for all conditions tested. Data points are labeled according to the (top) type of masker used and (bottom) processing method.

users, at least the variability obtained psychophysically and readily available from their MAPs. This is evident in Table V illustrating the different SRT values obtained by three subjects when tested in the same processing method and noise condition (Method B, SSN 90°). Subject's SRT values varied from a low of -11.5 dB to a high of -6 dB. The subjects' MAPs differed in the stimulation rate and dynamic range (the individual subject's thresholds and comfortable levels were also different). It is unclear whether differences in the values of these two parameters alone contributed to differences in SRT values as both parameters have been shown to have a weak effect on performance, at least in quiet (Vandali *et al.*, 2000; Fu and Shannon, 1999; Zeng and Galvin, 1999). The study by Zeng and Galvin (1999), for instance, has shown that a reduced dynamic range can severely degrade phoneme recognition in noisy conditions but not in quiet conditions. The three subjects (Table V) had different dynamic ranges with one having extremely low dynamic range (2–3 dB) suggesting that the dynamic range, along with other neuropathological factors, could have contributed to the differences in SRT values. The predicted SRT values (Table V) were quite close to the observed SRT values, thus demonstrating that the ECM model is capable of capturing the individual user's performance in different conditions. This is also confirmed with the individual correlations obtained for each subject for the 18 conditions tested (see Table IV). With the exception of two subjects, all individual correlations were quite high ($r > 0.91$).

Relatively good agreement was noted for both steady and speech maskers. This finding stands in contrast with the inability of most intelligibility measures (evaluated with NH listeners) to predict speech intelligibility in speech maskers (Rhebergen and Versfeld, 2005; Festen and Plomp, 1990). The good agreement observed in the present study with the intelligibility of speech maskers is due to the fact that CI users do not benefit from “dip listening” or glimpsing (Stickney *et al.*, 2004; Li and Loizou, 2010). Consequently, neither the model nor the CI users showed any advantage (masking release) with speech maskers. The correlation coefficients obtained with the three type of maskers (shown in the top panel of the figure), when considered separately, are 0.97, 0.92, and 0.96, respectively.

V. DISCUSSION AND CLINICAL IMPLICATIONS

Good correspondence ($r = 0.96$) was obtained between the SRTs obtained by CI users in noisy conditions and the

TABLE V. Observed and predicted SRT values for three subjects tested in the same method and noise condition (Method B, SSN 90°)

Subject	MAP parameters			SRT observed (dB)	SRT predicted (dB)
	Number of active electrodes	Stimulation rate (Hz)	Mean dynamic range (L/R) (dB)		
S1	22	800	8.71/8.34	-11.5	-12
S3	22	889	2.57/2.47	-9.5	-10.79
S6	22	1231	5.01/7.25	-6	-8.76

predicted SRTs obtained by the proposed intelligibility model using solely information contained in the individual CI user's MAP (e.g., number of active electrodes, dynamic range, etc.). This was surprising at first given that other factors (e.g., Blamey *et al.*, 1996) that could potentially influence performance were not taken into account or built in the model. Among others, these factors include electrode insertion depth, insertion trauma, distance of electrodes from the neurons and number and pattern of surviving ganglion cells. To some extent, the psychophysical thresholds capture a wealth of information about the above factors (Pfungst *et al.*, 2004). High thresholds are observed, for instance, when the electrodes are located far from excitable neural elements or when the number of ganglion cells at a particular location adjacent to the electrodes is scarce (Pfungst *et al.*, 2004). Both threshold and comfortable values vary across the electrode array and across subjects. The variation in threshold values, in particular, has been shown by Pfingst *et al.* (2004) to correlate highly ($r = -0.7$) with consonant recognition scores (Pfungst *et al.*, 2004). In the context of our model, the variation in threshold and comfortable levels is captured by the compression function, which maps the acoustic amplitudes to electrical amplitudes. This mapping in turn alters the resulting output envelopes used in the computation of the ECM index [Eq. (1)] and possibly changes across-channel cues. Finally, it should be pointed out that in the present study the ECM model was trained and tested under a small set of listening conditions. Evaluation of the proposed ECM model in various noisy and reverberant environments will be the main directions of future research.

Clearly, the ECM index can provide useful insights about the contribution and interaction between fitting parameters and speech recognition. But, can the proposed ECM index also be used clinically? Can the audiologists use it say to adjust and/or optimize the fitting parameters of individual CI users in order to improve their performance in noise? The answer is yes, and below we summarize two ways that the ECM index can be used in a clinical setting. The limited amount of time (1–2 h) the audiologists can afford to spend with CI users in the clinic prevents them from trying out different parameters in different backgrounds and finding out which works the best. The ECM index offers a good alternative which requires minimum amount of time to execute. Audiologists can use the ECM index in the following two ways.

1. Optimize choice of fitting parameters for individual CI users. There are number of fitting parameters that the audiologist can change using the clinical fitting software, and these include, among others, number of active electrodes and shape of compression function (Q). The shape of the compression function has been found to influence performance in noisy conditions (Fu and Shannon, 2000). In particular, a less compressive function was shown by Fu and Shannon (2000) to produce higher vowel and consonant recognition scores in noise, compared to the use of a highly compressive function. Audiologists could thus run the ECM model for a range of Q values in different noise conditions (and at different SNR levels) often encountered by the user, and pick the value of Q that yields the lowest SRT value. A

new “noise” MAP can subsequently be programmed with this new value of Q and given to the CI user to take home to use in noisy conditions.

2. Choose the right “noise” program for individual CI users. The present study evaluated the performance of three different noise reduction algorithms (Methods A, B and C) in various noise conditions involving speech-weighted noise and competing talker(s) as maskers. The listening test for each subject lasted for about 1–2 days. Clearly, this is an unrealistically long task which could not be done clinically to determine which of the three noise-reduction methods would work the best for a particular CI user. Alternatively, the audiologist could run the speech processed by each of the three methods through the ECM model, and pick the method that yields the lowest SRT in most conditions. Method C was found in the present study to yield the lowest SRT in most conditions, consistent with the predictions of the ECM model (see Fig. 3). Audiologists would thus program the user's processor with Method C and designate that as the “noise” program to take home.

The clinicians are not the only ones that could benefit from the use of the ECM model. The engineers who design noise-reduction algorithms could also benefit a great deal. Most noise-reduction algorithms are dependent on a number of parameters that often require tuning. Scanning the parametric space for the optimal parameter by testing CI users in different noise environments is an enormous and time-consuming task. The engineers could instead run the ECM model with the different set of parameters and obtain the optimal parameter in a matter of hours. The ECM model can therefore be used by design engineers to develop new and better noise reduction algorithms in a more timely fashion.

The proposed ECM model was designed to predict the performance (in terms of SRT) of individual CI users (or group of users) in noisy conditions following a change in the user's MAP (e.g., a change in the number of active electrodes). It can also be used by audiologists to select a noise-reduction algorithm that would work the best for a particular CI user. The ECM model was not designed to predict the errors made in a phoneme identification task (e.g., vowel recognition). For such a task, the multidimensional phoneme identification (MPI) model proposed by Svirsky and colleagues (Svirsky, 2000; Svirsky *et al.*, 2011), would be more appropriate. The MPI model has been found to reliably predict the error patterns produced by CI users in vowel and consonant identification tasks (Svirsky *et al.*, 2011). In brief, the ECM and MPI models serve two different purposes, with the latter being used primarily to understand how CI users can identify vowels or consonants given the degraded temporal/spectral cues they have at their disposal.

VI. CONCLUSIONS

High correlation ($r = 0.96$) was observed between the SRT values obtained by CI users and the output of the proposed envelope-based correlation measure (ECM). The findings from the present study suggest that the information present in electrograms is sufficient to predict speech

intelligibility by CI users. The proposed modulation-based ECM index makes use of low-frequency ($f < 25$ Hz) envelope modulations to make its prediction. High correlation was obtained with the ECM index despite the fact that other neuropathological factors (e.g., distance of electrodes to excitable neurons) known to influence performance (and difficult to measure) were not incorporated in the model. We believe that the psychophysical detection thresholds contain important information about these factors (Pfungst *et al.*, 2004). The compression function uses the thresholds of individual electrodes in order to map the acoustic to electrical envelopes. Changes in the electrical envelopes, caused by the variation in thresholds and comfortable levels across the electrode array, influence the covariance between the probe and response envelopes [i.e., value of r_k in Eq. (1)], and consequently the ECM value. The proposed ECM index captures the variability in SRT performance among CI users, and the predicted SRT values were in agreement with the observed ones. The individual correlations obtained for each CI user between observed and predicted SRTs were quite high ($r > 0.91$) for nearly all users. Some of the main advantages of the proposed ECM model include its clinical implications, simplicity, ease of use and the fact that it can be applied to individual CI users or a group of users, if the intent is to predict average (across users) performance. To use the model, audiologists only require the CI user's MAP. Finally, the model is also useful to engineers developing new noise reduction algorithms or optimizing/refining algorithmic parameters, a laborious task.

ACKNOWLEDGMENTS

This work was supported by Grant R01 DC 010494 awarded from NIDCD/NIH.

¹The probe is the envelope of clean speech (noise free) obtained after ACE implant processing and the response is the envelope of the processed (by one of the 3 noise reduction processing methods) speech obtained after ACE implant processing.

- ANSI (1997). S3.5-1997, *Methods for Calculation of the Speech Intelligibility Index* (American National Standards Institute, New York).
- Berge, J. V., and Wouters, J. (1998). "An adaptive noise canceller for hearing aids using two nearby microphones," *J. Acoust. Soc. Am.* **103**, 3621–3626.
- Blamey, P., Arndt, P., and Bergeron, F. (1996). "Factors affecting auditory performance of postlingually deaf adults using cochlear implants," *Audiol. Neurotol.* **9**, 293–306.
- Chen, F., and Loizou, P. C. (2011). "Predicting the intelligibility of vocoded speech," *Ear Hear.* **32**, 331–338.
- Drullman, R., Festen, J. M., and Plomp, R. (1994). "Effect of reducing slow temporal modulations on speech reception," *J. Acoust. Soc. Am.* **95**, 2670–2680.
- Elliott, T. M., and Theunissen, F. E. (2009). "The modulation transfer function for speech intelligibility" *PLoS Comput. Biol.* **5**, 1–14.
- Festen, J., and Plomp, R. (1990). "Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing," *J. Acoust. Soc. Am.* **88**, 1725–1736.
- Fletcher, H., and Galt, R. H. (1950). "The perception of speech and its relation to telephony," *J. Acoust. Soc. Am.* **22**, 89–151.
- Friesen, L. M., Shannon, R. V., Baskent, D., and Wang, X. (2001). "Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants," *J. Acoust. Soc. Am.* **110**, 1150–1163.
- Fu, Q. J., and Shannon, R. V. (1999). "Phoneme recognition by cochlear implant users as a function of signal-to-noise ratio and nonlinear amplitude mapping," *J. Acoust. Soc. Am.* **106**, 18–23.
- Fu, Q. J., and Shannon, R. V. (2000). "Effects of dynamic range and amplitude mapping on phoneme recognition in nucleus-22 cochlear implant users," *Ear Hear.* **21**, 227–235.
- Goldsworthy, R. L., and Greenberg, J. E. (2004). "Analysis of speech-based speech transmission index methods with implications for nonlinear operations," *J. Acoust. Soc. Am.* **116**, 3679–3689.
- Hawley, M. L., Litovsky, R. Y., and Culling, J. F. (2004). "The benefit of binaural hearing in a cocktail party: Effect of location and type of interferer," *J. Acoust. Soc. Am.* **115**, 883–843.
- Holube, I., and Kollmeier, B. (1996). "Speech intelligibility prediction in hearing-impaired listeners based on a psychoacoustically motivated perception model," *J. Acoust. Soc. Am.* **100**, 1703–1716.
- Houtgast, T., and Steeneken, H. J. M. (1985). "A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria," *J. Acoust. Soc. Am.* **77**, 1069–1077.
- IEEE (1969). "IEEE recommended practice for speech quality measurements," *IEEE Trans. Audio Electroacoust.* **AU-17**, 225–246.
- Levitt, H. (1971). "Transformed up-down methods in psychoacoustics," *J. Acoust. Soc. Am.* **49**, 467–477.
- Li, N., and Loizou, P. C. (2010). "Masking release and the contribution of obstruent consonants on speech recognition in noise by cochlear implant users," *J. Acoust. Soc. Am.* **129**, 1262–1271.
- Ma, J., Hu, Y., and Loizou, P. C. (2009). "Objective measures for predicting speech intelligibility in noisy conditions based on new band-importance functions," *J. Acoust. Soc. Am.* **125**, 3387–3405.
- Pfungst, B. E., Xu, L., and Thompson, C. S. (2004). "Across-site threshold variation in cochlear implants: Relation to speech recognition," *Audiol. Neurotol.* **9**, 341–352.
- Rhebergen, K. S., and Versfeld, N. J. (2005). "A speech intelligibility index-based approach to predict the speech reception threshold for sentences in fluctuating noise for normal-hearing listeners," *J. Acoust. Soc. Am.* **117**, 2181–2192.
- Spahr, A. J., Dorman, M. F., Litvak, L. M., Van Wie, S., Gifford, R. H., Loizou, P. C., Loisel, L. M., Oakes, T., and Cook, S. (2012). "Development and validation of the AzBio sentence lists," *Ear Hear.* **33**, 112–117.
- Stickney, G. S., Zeng, F. G., Litovsky, R., and Assmann, P. (2004). "Cochlear implant speech recognition with speech maskers," *J. Acoust. Soc. Am.* **116**, 1081–1091.
- Svirsky, M. A. (2000). "Mathematical modeling of vowel perception by users of analog multichannel cochlear implants: Temporal and channel-amplitude cues," *J. Acoust. Soc. Am.* **107**, 1521–1529.
- Svirsky, M. A., Sagi, E., Meyer, T. A., Kaiser, A. R., and Teoh, S. W. (2011). "A mathematical model of medial consonant identification by cochlear implant users," *J. Acoust. Soc. Am.* **129**, 2191–2200.
- Vandali, A. E., Whitford, L. A., Plant, K. L., and Clark, G. M. (2000). "Speech perception as a function of electrical stimulation rate: Using the nucleus 24 cochlear implant system," *Ear Hear.* **21**, 608–624.
- Wójcicki, K. K., and Loizou, P. C. (2012). "Channel selection in the modulation domain for improved speech intelligibility in noise," *J. Acoust. Soc. Am.* **131**, 2904–2913.
- Yousefian, N., and Loizou, P. C. (2011). "A dual-microphone algorithm for improving speech intelligibility in competing-talker situations," in *Proc. Conference on Implantable Auditory Prostheses (CIAP'11)* (Pacific Grove, CA).
- Yousefian, N., and Loizou, P. C. (2012). "A dual-microphone speech enhancement algorithm based on the coherence function," *IEEE Trans. Speech Audio Process.* **20**, 599–609.
- Zeng, F., and Galvin, J., III. (1999). "Amplitude mapping and phoneme recognition in cochlear implant listeners," *Ear Hear.* **20**, 60–74.